Learning the reasons why groups of consumers prefer some food products

Juan José del Coz[1], Jorge Díez[1], Antonio Bahamonde[1], Carlos Sañudo[2], Matilde Alfonso[2], Philippe Berge[3], Eric Dransfield[3], Costas Stamataris[4], Demetrios Zygoyiannis[4], Tyri Valdimarsdottir[5], Edi Piasentier[6], Geoffrey Nute[7], Alan Fisher[7]

[1] Artificial Intelligence Center, University of Oviedo at Gijón, 33204 Gijón Spain
{juanjo, jdiez, antonio}@aic.uniovi.es
[2] Facultad de Veterinaria. University of Zaragoza, Zaragoza (Aragón), Spain
csanudo@unizar.es, matilde_alfonso@ncsu.edu
[3] Unité de Recherches sur la Viande, INRA de Theeix, 63100 St. Genès- Champanelle, France. Current address: Wageningen UR, 6700 AA wageningen, The Netherlands
{berge, dransf }@clermont.inra.fr
[4] Aristotle University, Department of Animal Health and Husbandry, Thessaloniki, Greece
{stamatar, zygoyan}@vet.auth.gr
[5] Icelandic Fisheries Laboratories, PO Box 1405, 121, Reykjavík, Iceland
maturoglyf@simnet.is
[6] Department de Science della Produzione Animale, University of Udinem, Pagnacco, Italy.
edi.piasentier@uniud.it
[7] Department of Food Animal Science, University of Bristol, BS40 5DU, United Kingdom
{geoff.nute, alan.fisher}@bris.ac.uk

Abstract. In this paper we propose a method for learning the reasons why groups of consumers prefer some food products instead of others of the same type. We emphasize the role of groups given that, from a practical point of view, they may represent market segments that demand different products. Our method starts representing in a metric space people preferences; there we are able to define similarity functions that allow a clustering algorithm to discover significant groups of consumers with homogeneous tastes. Finally in each cluster, we learn, with a SVM, a function that explains the tastes of the consumers grouped in the cluster. Additionally, a feature selection process highlights the essential properties of food products that have a major influence on their acceptability. To illustrate our method, a real case of consumers of lamb meat was studied. The panel was formed by 773 people of 216 families from 6 European countries. Different tastes between Northern and Southern families were enhanced.

1 Introduction

Consumer preferences for food products address the strategies of industries and breeders, and should be carefully considered when export and commercial policies are designed. In this paper we present a method to deal with data collected from panels of consumers in order to discover groups with differentiated tastes; these groups may constitute significant market segments that demand different kinds of food products. Additionally, our approach studies the factors that contribute to the success or failure of food products in each segment.

From a conceptual point of view, the panels are made up of untrained consumers; these are asked to rate their degree of acceptance or satisfaction about the tested products on a scale. The aim is to be able to relate product descriptions (human and mechanical) with consumer preferences. Simple statistical methods can not cope with this task. In fact, this is not a straightforward task; the reason is that when we are aiming to induce a function that maps object descriptions into ratings, we must consider that consumers' ratings are just a way to express their preferences about the products presented in the same testing session. Additionally, it is necessary to realize that numerical ratings do not mean the same for all the people, the scales used may be quite different. Discussions about ratings and preferences can be found in [1], in the context of food preferences in [2, 3, 4].

To illustrate our method, we used a data set that collects the ratings of a panel of lamb meat consumers. Let us recall that the world market for this meat is quite important; in fact, among all meats, lamb meat is the most internationally traded, 15% of total world production is exported.

The panel studied was formed by 216 European families, from 6 countries, that ordered, according to their preferences, 12 kinds of lambs [5, 6]. The purpose of the study was to discover what features of lambs may explain why significant groups of consumers prefer some lamb types. Thus, we start looking for significant *clusters* of families with similar tastes.

The main assumption behind the approach presented in this paper is that we are able to map people's preferences into a metric space in such a way that we can assume some kind of continuity. In the case of lamb meat panel, the mapping can be simply given by a ranking vector of lamb types provided by each consumer or family of consumers.
However, this is not the general case. Thus, we extended the method proposed here to situations where the size of the sample of food prevents panellist from testing all products. We must take into account that usually we can not ask our panellist to spend long periods of time rating the whole set of food samples. Typically, each consumer only participates in one or a small number of testing sessions, usually in the same day. Notice that tasting a large sample of food may result physically impossible, or the number of tests performed would damage the sensory capacity of consumers. In this case we will codify people preferences by the weighting vector of a linear function in a high dimensional space; the space where we represent the descriptions of food products. Thus, the similarity is defined by means of the kernel attached to the representation map. This approach has been successfully applied in [7].

Once we have people's preferences represented in a metric space, and we have defined a similarity function, then we use a clustering algorithm. Although there are other possibilities, we used the nonparametric hierarchical clustering algorithm of Dubnov *et al*. [8] that uses a proximity matrix of pairwise relations that directly captures the intention of the similarity functions. Then in each cluster, we learn a ranking function from the descriptions of each object involved in testing sessions; so we will be able to explain why the group of consumers of the cluster prefers some kind of products instead of others. Moreover, a feature selection algorithm will point out the essential characteristics that make the difference between success and failure in the market segment that clusters represent.

The paper is organized as follows. In the next section we describe how it is possible to measure similarities between preference criteria of two consumers. In the third section we explain the clustering algorithm used. The last section is devoted to report the results achieved in the case of the data from the panel of European lamb meat consumers. We spell out the steps followed by our method in this real world case, and we review the implications both to lamb breeders and to designers of future commercial strategies.

2 Computing distances between preference criteria

This section is devoted to show how preference criteria of consumers can be mapped into a metric space where it is possible to define a similarity measure. We distinguish two situations. In the first one, each consumer already provides a ranking vector, while we require a kernel based method in the most general case.

2.1 When everybody tastes everything: using explicit rankings

In some cases, we have situations where there are a fixed set of food items that each tester can order in a ranking of preferences. Then, the similarity of tester preferences can be straightforward measured. In this section we analyze one of these cases, the data collected on a large panel of European families testing 12 different types of lambs.

Both lambs and families were selected from 6 different countries: Greece, Italy, and Spain (Southern countries in the European Union), and France, Iceland, and United Kingdom (Northern countries). A total of 36 families in each country rated each lamb sample (a total of 216 families); we considered the average rating of each family as the *unit* expression of their preferences; a total of 773 people were involved in the panel. The decision of averaging the ratings into each family is justified on [5], where it was noticed that there is more agreement in the rates between individuals within a household than between households; that means that there exists an important effect that might be called *family halo* in people’s gastronomic preferences.

The panel was asked to rate several aspects of lamb meat on a numerical scale from 0 to 100; however, we are only going to deal with their *overall* judgement. Testing was done over a period between 3 and 6 months depending of the country. Each family received 12 hind leg joints (one from each lamb type), and they were asked to roast the joints using their own cooking criteria. It is important to notice that 108 lambs per type were used, what means 1296 animals, and 2592 hind legs. The sample is a quite wide range of lamb covering different breeds, diets, age at slaughter and weights of carcass; see Table 1 for more details.

Table 1. Description of lamb types in order of increasing carcass weight in Kg. The average age at slaughter is expressed in months

Country of origin	Breed type	Age at slaughter	Carcass Weight	Main feeding background	Lamb Codes
Spain (ES)	Churra	1.0	5.4	Milk	4
Greece (GR)	Karagouniko	1.7	8.1	Milk	7
Spain (ES)	Rasa Aragonesa	2.8	10.0	Concentrate	3
Italy (IT)	Appenninica	2.4	11.2	Concentrate	12
United Kingdom (GB)	Welsh Mountain	7.4	15.3	Grass	2
France (FR)	Lacaune	3.3	15.3	Concentrate	6
Greece (GR)	Karagouniko	3.5	15.4	Concentrate	8
Iceland (IS)	Icelandic	4.3	15.9	Grass	10
France (FR)	Meat breeds	7.0	16.6	Grass	5
Iceland (IS)	Icelandic	4.3	16.7	Grass	9
United Kingdom (GB)	Suffolk x Mule	4.0	17.8	Grass	1
Italy (IT)	Bergamasca	12.0	30.5	Transhumance	11

The preferences expressed by each family were summarized by the ranking of lamb types ordered according to their rates. Then, the similarity of the preferences of two families was computed as the number of pairs where both rankings coincide in their relative ordering; in this case, an integer from 0 to 66. In symbols, if r_1 and r_2 are two rankings, we define

$$similarity(r_1, r_2) = \sum_{t_1, t_2 \in LT, t_1 \neq t_2} \mathbf{1}_{((r_1(t_1) - r_1(t_2)) * (r_2(t_1) - r_2(t_2)) \geq 0)} \qquad \textbf{(1)}$$

where LT is the set of lamb types; $\mathbf{1}_{(p(x))}$ returns 1 when p(x) is true and 0 otherwise; and $r_i(t_j)$ stands for the ordinal number of lamb type t_j in ranking r_i.

2.2 In a general case: using ranking functions

In this section we deal with a more general case (see [7]) than that of lambs spelled out in the previous section. Now we assume that the consumers involved in a panel can be divided into sociological categories or *units*, and that each person has rated a limited number of samples in one or a few sessions. Therefore it is not straightforward to compute a ranking of food products for each unit. Instead of that, we are going to induce a function able to captures somehow the criteria used to express unit preferences. Then we will manage to define similarities in the space of those functions.

Although there are other approaches to learn preferences, following [9, 10, 11] we will try to induce a real *preference*, *ranking*, or *utility function* f from the space of object descriptions, say $\mathbf{R}^d$, in such a way that it maximizes the probability of having $f(\mathbf{x}) > f(\mathbf{y})$ whenever $\mathbf{x}$ is preferable to $\mathbf{y}$; we call such pairs, *preference judgments*. This functional approach can start from a set of objects endowed with a (usually ordinal) rating, as in regression; but essentially, we only need a collection of preference judgments.

When we have a set of ratings given by members of a unit u, we most take into account the session where the ratings have been assessed [2, 4]. Thus, for each session we consider the average of all ratings given by members of the unit to each sample presented in the session; then we include in the set of preference judgments PJ_u the pairs ($\mathbf{x}$, $\mathbf{y}$) whenever the sample represented by $\mathbf{x}$ had higher rating than the sample represented by $\mathbf{y}$. In this way, we can overcome the *batch effect*: a product will obtain a higher/lower rating when it is assessed together with other products that are clearly worse/better. In fact, if we try to deal with sensory data as a regression problem, we will fail [3]; due to the batch effect, the ratings have no numerical meaning: they are only a relative way to express preferences between products of the same session.

In order to induce the ranking function, we can use the approach presented by Herbrich *et al*. in [9]. So, we look for a function F_u: $\mathbf{R^d} \times \mathbf{R^d} \to \mathbf{R}$ such that

$$\forall \mathbf{x},\mathbf{y} \in \mathbf{R}^d,\ F_u(\mathbf{x},\mathbf{y}) > 0 \Leftrightarrow F_u(\mathbf{x},\mathbf{0}) > F_u(\mathbf{y},\mathbf{0}) \qquad (2)$$

Then, the ranking function f_u: $\mathbf{R^d} \to \mathbf{R}$ can be defined by $F_u(\mathbf{x},\mathbf{0})$ plus any constant.

Given the set of preference judgments PJ_u, we can specify F_u by means of the constraints

$$\forall\ (\mathbf{x}, \mathbf{y}) \in PJ_u,\ F_u(\mathbf{x}, \mathbf{y}) > 0 \text{ and } F_u(\mathbf{y}, \mathbf{x}) < 0 \qquad (3)$$

Therefore, PJ_u gives rise to a set of binary classification training set

$$E_u = \{(\mathbf{x}, \mathbf{y}, +1), (\mathbf{y}, \mathbf{x}, -1): (\mathbf{x}, \mathbf{y}) \in PJ_u\} \qquad (4)$$

Nevertheless, a separating function for E_u does not necessarily fulfill (2). Thus, we need an additional constraint about the antisymmetrical role that we require for the objects of E_u entries. So, if we represent each object description $\mathbf{x}$ in a higher dimensional feature space by means of $\phi(\mathbf{x})$, then we can represent pairs ($\mathbf{x}$, $\mathbf{y}$) by $\phi(\mathbf{x})$ - $\phi(\mathbf{y})$. Hence, a classification SVM can induce the function of the form:

$$F_u(\mathbf{x},\mathbf{y}) = \sum_{s \in SV_u} \alpha_s z_s \left\langle \phi(\mathbf{x}_s^{(1)}) - \phi(\mathbf{x}_s^{(2)}), \phi(\mathbf{x}) - \phi(\mathbf{y}) \right\rangle \qquad (5)$$

where $<\mathbf{x}, \mathbf{y}>$ stands for the inner product of vectors $\mathbf{x}$ and $\mathbf{y}$; SV_u is the set of support vectors, notice that they are formed by two d-dimensional vectors $(\mathbf{x}_s^{(1)}, \mathbf{x}_s^{(2)})$, while the scalars z_s represent the class $+1$ or -1. Trivially, F_u fulfils the condition (2).

Notice that if k is a kernel function defined as the inner product of two objects represented in the feature space, that is, $k(\mathbf{x}, \mathbf{y}) = <\phi(\mathbf{x}), \phi(\mathbf{y})>$, then the kernel function used to induce F_u is

$$\mathbf{K}(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \mathbf{x}_4) = k(\mathbf{x}_1, \mathbf{x}_3) - k(\mathbf{x}_1, \mathbf{x}_4) - k(\mathbf{x}_2, \mathbf{x}_3) + k(\mathbf{x}_2, \mathbf{x}_4) \qquad (6)$$

Usually it is employed a linear or a simple polynomial kernel; that is, $k(\mathbf{x}, \mathbf{y}) = \langle \mathbf{x}, \mathbf{y} \rangle$, or $k(\mathbf{x}, \mathbf{y}) = (\langle \mathbf{x}, \mathbf{y} \rangle + c)^g$, with $c = 1$ and $g = 2$.

Once we have a function F_u for a unit u fulfilling (2), then a utility function f_u is given by

$$f_u(x) = \sum_{s \in SV_u} \alpha_s z_s \left\langle \phi(\mathbf{x}_s^{(1)}) - \phi(\mathbf{x}_s^{(2)}), \phi(\mathbf{x}) \right\rangle = \sum_{s \in SV_u} \alpha_s z_s \left(k(\mathbf{x}_s^{(1)}, \mathbf{x}) - k(\mathbf{x}_s^{(2)}, \mathbf{x}) \right) \qquad (7)$$

Therefore, f_u can be represented by the weight vector $\mathbf{w}_u$ in the higher dimensional space of features such that

$$f_u(x) = \langle \mathbf{w}_u, \phi(\mathbf{x}) \rangle, \tag{8}$$

where

$$\mathbf{w}_u = \sum_{s \in SV_u} \alpha_s z_s \left(\phi(\mathbf{x}_s^{(1)}) - \phi(\mathbf{x}_s^{(2)}) \right) \tag{9}$$

Now we only need to define the distance of unit preferences. Given that preferences are codified by those weighting vectors, we define the similarity of the preferences of units u and u′ by the cosine of their weighting vectors. In symbols,

$$similarity(\mathbf{w}_u, \mathbf{w}_{u'}) = \cos(\mathbf{w}_u, \mathbf{w}_{u'}) = \frac{\langle \mathbf{w}_u, \mathbf{w}_{u'} \rangle}{\|\mathbf{w}_u\| * \|\mathbf{w}_{u'}\|} \tag{10}$$

Given that this definition uses scalar products instead of coordinates of weighting vectors, we can easily rewrite (10) in terms of the kernels used in the previous derivations. The essential equality is:

$$\begin{aligned} \langle \mathbf{w}_u, \mathbf{w}_{u'} \rangle &= \sum_{s \in SV_u} \sum_{l \in SV_{u'}} \alpha_s \alpha_l z_s z_l \left\langle \phi(\mathbf{x}_s^{(1)}) - \phi(\mathbf{x}_s^{(2)}), \phi(\mathbf{x}_l^{(1)}) - \phi(\mathbf{x}_l^{(2)}) \right\rangle \\ &= \sum_{s \in SV_u} \sum_{l \in SV_{u'}} \alpha_s \alpha_l z_s z_l \mathbf{K}\left(\mathbf{x}_s^{(1)}, \mathbf{x}_s^{(2)}, \mathbf{x}_l^{(1)}, \mathbf{x}_l^{(2)} \right) \end{aligned} \tag{11}$$

3 Generalizing preferences from consumers to groups

Once we have defined a reasonable similarity measure for preference criteria, we proceed to look for clusters of consumers with homogeneous tastes. In principle, we could use any available clustering algorithm. However, we avoided those methods, like k-means, that require frequent recomputations of the centroids of each cluster. The reason is that the updating of (11) would result very uncomfortable. Additionally, we need a mechanism able to estimate a reasonably number of clusters directly from the data, without any explicit manual intervention.

Hence, we applied a nonparametric pairwise algorithm of Dubnov *et al.* [8], although this is not probably the only possibility. The following paragraphs sketch a description of this algorithm as we used it in the experimental results reported in the last section.

3.1 The clustering algorithm

Let $S = (s_{ij})$ be a square matrix where s_{ij} stands for the similarity between data points i and j; in our case, data points are the vectorial representation of the preference criteria of consumer units, and similarities are given by equations (1) or (10). In the following, S will be called the *proximity matrix*.

The matrix S is transformed iteratively, following a two step procedure that makes it to converge to a binary matrix, yielding a bipartition of the data set into two clusters. Then, recursively, the partition mechanism is applied to each of the resulting clusters represented by their corresponding submatrices. To guarantee that only meaningful splits take places, Dubnov *et al*. [8] provide a cross validation method that measures an index that can be read as a significance level; we will only accept splits in which the level is above 95%.

The basic iterative transformation uses the following formulae to go from iteration t to t+1:

$$p_{ij}(t+1) = \frac{s_{ij}(t)}{max\{|s_{ik}(t)| : k\}} \tag{12}$$

$$s_{ij}(t+1) = \frac{1}{2}\sum_k p_{ik}(t+1) \log \frac{p_{ik}(t+1)}{\frac{1}{2}\left(p_{ik}(t+1) + p_{jk}(t+1)\right)} + \frac{1}{2}\sum_k p_{jk}(t+1) \log \frac{p_{jk}(t+1)}{\frac{1}{2}\left(p_{jk}(t+1) + p_{ik}(t+1)\right)}$$

The first step gives rise to (p_{ij}) normalizing the columns of the proximity matrix using the L_∞ norm; then the proximities are re-estimated using the Jensen-Shannon divergence. The idea is to formalize that two preference criteria are close (after these two steps) if they were both similar and dissimilar to analogous sets of criteria before the transformation.

This method of clustering preference criteria is quite different from a work presented in [12]. That approach is based on the estimation of learning errors in the data sets of groups; therefore, the method requires a lot of data available, what make difficult its use when we are dealing with sensory data since the amount of data available is usually very scarce. Additionally, that method is a bottom-up clustering algorithm which tends to produce many clusters. In sensorial analysis applications, we don't expect that many market segments exist, so a top-down clustering is more adequate.

3.2 The preference function of groups

Given a set of clusters {Cluster(j): j = 1:n}, we have to explain the reasons that make people of each cluster to have those similar criteria that make them different from people of other clusters. The best way to achieve this is to induce a preference function using product descriptions. The learning algorithm is the SVM explained in section 2.2, but notice that now instead of using the preference judgments PJ_u sets of individual units, we consider for each cluster the union

$$PJ_{cluster(j)} = \bigcup_{u \in cluster(j)} PJ_u \tag{13}$$

The preference functions (see equation (7)) will be useful for two different things. First, we can compute the average ranking of the cluster, and the estimation of the ranking position of future products given their descriptions. Second, we can determine the influence of each feature that describes food products in the acceptability by consumers of the market segment represented by clusters. Therefore, we will be able to design policies to improve the acceptability by different kinds of consumers.

Feature influence analysis is not a straightforward task and it must be handled with care to obtain useful results. Different approaches must be used depending if we deal with linear or non-linear functions [13, 14, 15, 16]. For the aims of this paper, we use adaptations of these selection algorithms to preference learning [2, 4, 11].

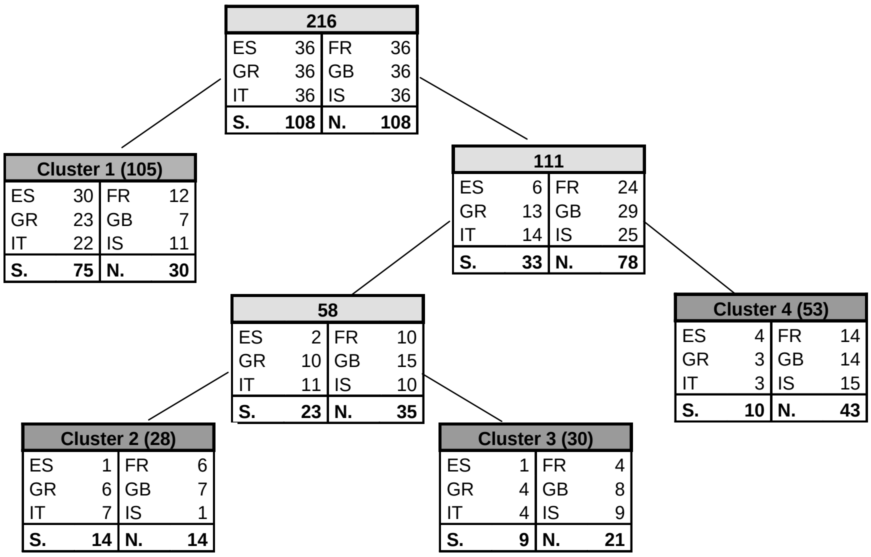

Fig. 1. Trace of the clustering algorithm. In each node we report the total number of families, the number of families of each country, and the sum of families from Southern and Northern European countries

4 Experimental results

In this section we report the results obtained with the data base of the European lamb panel. As was described in section 2.1, the distances between the preferences of two families was computed as the number of pairs with a disagreement in their relative order.

The clustering algorithm [8] returns the tree of 4 leaves depicted in Figure 1. All split nodes achieved a high confidence level: 100%, 95% and 97% respectively from top to bottom. Clusters 1 and 4 are the biggest; they sum 105 and 53 families, while the other two clusters represent minority market segments of 28 and 30 families each. The rankings of lamb types that capture the preferences of each cluster are reported at Table 2.

The degree of consensus achieved into the clusters can be estimated by the cross validation error of the classification SVM used to compute the rankings of each cluster when we merged the preference judgments of all families involved. In this case, each lamb type was described by 98 attributes reporting the data included in Table 1, chemical and physical properties of the meat, and a sensory description given by a set of trained experts. The cross validation estimations for these errors are, for cluster 1 to 4, 29.25%, 35.59%, 30.94%, and 36.64%, respectively. It is important to notice here that if, for each pair of lambs, we choose the most frequent relative ordering in each cluster, then the number of disagreements would be, 28.52%, 30.02%, 29.32%, and 32.31%, respectively in the four clusters. Therefore, the estimation of accuracy of the preferences functions induced is quite high.

Additionally, the processes of preference learning provide us the scores reported in Table 2; they are computed as the normalized values (in the range [0, 100]) returned by the corresponding ranking functions f_u (equation 7) of the clusters. These values can be interpreted as the *average* ratings into the clusters, but considering the individual ratings just as preference judgments instead of numeric values.

Table 2. . Composition and rankings for each cluster. We report the number of families from South and North European countries. The ranking of lamb types is shown by columns; cells shaded correspond to types with a country of origin in the South of Europe; the score column gives the normalized (between 0 and 100) outputs of the ranking function (learned for the corresponding cluster) in each lamb type

	Cluster 1		Cluster 2		Cluster 3		Cluster 4	
	# Families	%	# Families	%	# Families	%	# Families	%
South	75	71.4%	14	50%	9	30%	10	18.9%
North	30	28.6%	14	50%	21	70%	43	81.1%
Rank	**Lamb type**	**Score**	**Lamb type**	**Score**	**Lamb type**	**Score**	**Lamb type**	**Score**
1	4	100	11	100	11	100	6	100
2	3	84,5	5	62,8	8	72,8	5	89,3
3	12	77,5	1	51,1	2	68,9	8	89,3
4	6	64,9	4	47,1	6	63,6	9	84,7
5	7	62,7	3	46,4	10	58,5	10	70,5
6	8	58,1	2	37,6	3	55,9	1	68,2
7	9	55,4	8	28,7	1	46,8	2	67,8
8	5	51,0	9	18,5	9	46,4	7	59,4
9	1	45,0	6	10,4	5	40,8	12	59,1
10	2	44,6	7	3,5	7	34,5	3	25,8
11	10	44,0	12	1,4	12	30,0	4	23,4
12	11	0	10	0	4	0	11	0

4.1 Implications for lamb markets and breeders

In general, it is well known that meat qualities are mainly the result of a set of complex factors somehow inherent in animal’s breed, rearing system, and feeding background. With the panel available, we can try to quantify these biological complexities with the additional difficulty of measuring meat qualities through the sensorial appreciations of people with different geographical extractions.

In this sense, there are several conclusions that can be drawn from Table 2. First, we observe that the lamb type of code 11, the oldest and heaviest (see Table 1), divides the preferences of clusters; so while in clusters 1 and 4 this lamb type is the least appreciated, on the other two clusters, it is the most appreciated lamb type. This is a lamb with a very strong flavor and odor what arouses vivid reactions.

However, the most striking result is that the most representative clusters arrange the majority of families from Southern and Northern European countries respectively. Moreover, lamb types with origin in southern countries are the most appreciated in Southern countries, and the same happens if we refer to Northern countries and lambs. Most people like best the kind of lambs that they are used to eat at home. In other words, European lamb consumers seem to be very influenced by their culinary and cultural background.

To illustrate this point, we only have to observe the opposite role played by the sequence of lamb types 4, 3, and 12. While they occupy the leading positions in the mainly Southern cluster 1; they are relegated to the bottom of the list in the cluster of the mainly Northern families (clusters 4 and 3). These lamb types are the lightest (if we exclude the type 7) with a milk and concentrate diets.

Another important source of information is the relevancy of the features that take part in the learning process. In this case, the most relevant descriptors of lamb types in each cluster ranking are *phospholipids (php) fraction*, and *neural lipid (nl) fatty acids fraction*. However, from a practical point of view this information is not directly practicable; since it is not obvious at all how we can *improve* the *php* or the *nl* of lambs. Notice that the term ‘improve’ is a relative expression in this context, since its exact meaning depends on the cluster.

The question is what *visible* lamb features can be identified with people’s preferences? Or how can a breeder produce lambs for a given market segment? To answer these questions we have to realize that there are some features like age, weight, and feeding, that are easily modifiable by breeders. Moreover, using only these features and their products, it is possible to explain the rankings of each cluster. Thus, these features are not only visible and modifiable, but they content also enough information so as to approach a guide to breeders and designers of marketing strategies. Table 3 reports the contribution of these features to the ranking of each cluster.

Table 3. Contribution to the preferences in each cluster of the main attributes of lamb types: those where breeders can act over them in order to improve the popularity of their lamb meats

	Cluster 1		Cluster 2		Cluster 3		Cluster 4	
Attribute	**Sign**	**Relevancy**	**Sign**	**Relevancy**	**Sign**	**Relevancy**	**Sign**	**Relevancy**
milk	+	4	-	2	-	5	-	1
grass	-	1	+	5	+	3	+	3
concentrate	+	3	-	4	+	2	+	4
age	-	5	+	1	+	4	-	2
weight	-	2	+	3	+	1	+	5

To obtain these data, from the preference judgments expressed by the ranking of each cluster, we built one classification training set. Each lamb type (see Table 1) was described using 5 features: age, weight, and 3 binary features, one to describe if the feeding background was or not milk, another for grass, and a third one for concentrate. Then if t_1 was preferred to t_2, we included $t_1 - t_2$ with class +1, and $t_2 - t_1$ with class –1 in the corresponding training set. Notice that each of these training sets has 132 examples.

According to section 2.2, the coefficients of the hyperplane that separates the positive and negative classes are the weights of the features in the ranking function: a computational model of the preferences of the cluster. In Table 3 we report the signs of these coefficients. Notice that there is not any feature with the same sign in all clusters. On the other hand, we observe that the differences in sign from cluster 1 to the other 3 is the biggest of any single cluster; this would explain, from another point of view, the clustering distribution proposed by the algorithm of Dubnov *et al*. [8]. The split of clusters 2, 3 and 4 can even be conjectured from their sign distributions.

In addition to sign we include in Table 3, the order of relevancy of each of the features. Here we used a procedure based on the Gram-Schmidt orthogonalization, see [16]. The information provided by this ordering allows us to gain insight into the strength of the features of lamb types when they are used to explain the preferences of a cluster.

From cluster 1 and cluster 4, related with South and Northern families respectively, we observe as in the South young animals, with small slaughter weight, reared with milk or concentrate are preferred. Whereas in the North countries, grass or concentrate fed lambs, with high carcass weights, are the most appreciated. These results are related with the culinary habits of the consumers, as was pointed out in [17]; additionally, the results justify the idea that it is necessary to produce lamb meat taking into consideration the destination market.

5 Conclusions

In this paper we propose a method for learning the reasons why groups of consumers prefer some food products instead of others of the same type. To illustrate our method, a real case of consumers of lamb meat was studied and we pointed out some practical conclusions that can allow breeders to design policies to improve the acceptability of lamb meat by different market segments.

The proposed method stresses that it is possible to map with continuity people's preferences into a metric space, where it is possible to compute the similarity between preference criteria. In this context we distinguish two kinds of situations: i) all consumers rate all products, then the similarity is computed as the number of preference judgements pairs where their rankings coincide; and ii) each consumer only rates some products of the sample, then we codify their preferences by linear functions in a high dimensional space and compute the similarity between these ranking functions by means of a kernel based method.

Once we have a reasonable similarity measure for preference criteria, the main goal is to discover different groups of consumers (or market segments) and explain why consumers of each group prefer some kind of products. For this purpose, we use three learning tools: i) to group people with similar preferences, a hierarchical clustering algorithm that directly captures the intention of the similarity functions using a proximity matrix of pairwise relations; ii) a SVM algorithm to learn preferences functions using the descriptions of the products; and iii) a feature selection algorithm to point out the essential characteristics that make the difference between success and failure in the market segment that each cluster represents.

Acknowledgments

The research reported in this paper is supported in part under the grant TIN2005-08288 from the Spanish Ministerio de Educación y Ciencia.
We would like to thank: the authors of Spider [18], a MatLab toolbox that includes kernel based algorithms; and Thorsten Joachims [19] for his SVM^{light}. Those systems were used in the experiments reported in this paper.

References

1. Cohen, W., Shapire, R., Singer, Y.: Learning to order things. Journal of Artificial Intelligence Research, 10 (1999) 243–270.
2. Del Coz, J. J., Bayón, G. F., Díez, J., Luaces, O., Bahamonde, A., Sañudo, C.: Trait selection for assessing beef meat quality using non-linear SVM. Proceedings of the Eighteenth Annual Conference on Neural Information Processing Systems (NIPS 2004). Vancouver, British Columbia, Canada, December (2004) 13–18.
3. Díez, J., Bayón, G. F., Quevedo, J. R., del Coz, J. J., Luaces, O., Alonso, J., Bahamonde, A.: Discovering relevancies in very difficult regression problems: applications to sensory data analysis. Proceedings of the European Conference on Artificial Intelligence (ECAI ’04), Valencia, Spain (2004) 993–994.
4. Luaces, O., Bayón, G.F., Quevedo, J.R., Díez, J., del Coz, J.J., Bahamonde, A.: Analyzing sensory data using non-linear preference learning with feature subset selection. Proceedings of the 15th European Conference of Machine Learning (2004) 286–297.
5. Dransfield, E., Martin, J-F., Fisher, A., Nute, G.R., Zygyiannis, D., Stamataris, C., Thorkelsson, G., Valdimarsdottir, T., Piasentier, E., Mills, C., Sañudo, C., Alfonso, M.: Home Placement Testing of Lamb Conducted in Six Countries. Journal of Sensory Studies, 15 (2000) : 421–436.
6. Sañudo, C., Alfonso, M., Sanchez, A., Berge, F., Dransfield, E., Zygoyiannis, D., Stamataris, C., Thorkelsson, G., Valdimarsdottir, T., Piasentier, E., Mills, C., Nute, G., Fisher, A.: Meat texture of lambs from different European production systems. Australian Journal of Agricultural Research, 54 (2003) 551–560.
7. Díez, J., del Coz, J. J., Sañudo, C., Albertí, P., Bahamonde, A.: A Kernel Based Method for Discovering Market Segments in Beef Meat. Proceedings of the 9th European Conference on Principles and Practice of Knowledge Discovery in Databases, (2005) 462–469.
8. Dubnov, S., El-Yaniv, R., Gdalyahu, Y., Schneidman, E., Tishby, N., Yona, G.: A New Nonparametric Pairwise Clustering Algorithm Based on Iterative Estimation of Distance Profiles. Machine Learning, 47 (2002) 35–61.
9. Herbrich, R., Graepel, T., Obermayer, K.: Large margin rank boundaries for ordinal regression. In A. Smola, P. Bartlett, B. Scholkopf, and D. Schuurmans, editors, Advances in Large Margin Classifiers, MIT Press, Cambridge, MA. (2000) 115–132.
10. Joachims, T.: Optimizing search engines using clickthrough data. In: Proceedings of the ACM Conference on Knowledge Discovery and Data Mining (KDD) (2002).

11. Bahamonde, A., Bayón, G. F., Díez, J., Quevedo, J. R., Luaces, O., del Coz, J. J., Alonso, J., Goyache, F.: Feature subset selection for learning preferences: a case study. Proceedings of the 21st International Conference on Machine Learning, ICML (2004) 49–56.
12. Díez, J., del Coz, J.J., Luaces, O., Bahamonde, A.: A clustering algorithm to find groups with homogeneous preferences. In the 26th Annual International ACM Conference on Research and Development in Information Retrieval (SIGIR '03), Workshop on Implicit Measures of User Interests and Preferences, Toronto, Canada, (2003).
13. Guyon, I., Weston, J., Barnhill, S., & Vapnik, V.: Gene selection for cancer classification using support vector machines. Machine Learning, 46 (2002) 389–422.
14. Rakotomamonjy, A.: Variable selection using SVM-based criteria. Journal of Machine Learning Research, 3 (2003) 1357–1370.
15. Degroeve, S., De Baets, B., Van de Peer, Y., Rouzé, P.: Feature subset selection for splice site prediction. Bioinformatics, 18 (2002) 75–83.
16. Stoppiglia, H., Dreyfus, G., Dubois, R., Oussar, Y.: Ranking a Random Feature for Variable and Feature Selection. Journal of Machine Learning Research; 3 (2003) 1399–1414.
17. Sañudo C, Nute GR, Campo MM, María GA, Baker A, Sierra I, Enser M, Wood JD. Assessment of commercial lamb meat quality by British ad Spanish taste panels. Meat Science 48 (1998) 91–100.
18. Weston, J., Elisseeff, A., BakIr, G., Sinz, F.: SPIDER: object-orientated machine learning library. http://www.kyb.tuebingen.mpg.de/bs/people/spider/.
19. Joachims, T.. Making large-Scale SVM Learning Practical. Advances in Kernel Methods - Support Vector Learning, B. Schölkopf and C. Burges and A. Smola (ed.), MIT-Press, (1999).